



Association Francophone de la Communication Parlée

Siège social : Université d'Avignon
339 chemin des Meinajaries - BP 1228
84911 AVIGNON CEDEX 9

Réf : LE-2002-5

Communiqué de l'AFCP¹ du 3 Décembre 2002 concernant l'identification des individus par leur voix

L'actualité récente amène l'Association Francophone de la Communication Parlée (AFCP) à réaffirmer la position collective de la communauté scientifique francophone sur la question de l'identification vocale des individus.

Un nécessaire devoir de précaution

Les scientifiques francophones spécialistes de l'étude de la voix, représentés initialement par le Groupe Francophone de la Communication Parlée, puis depuis 2001 par l'Association Francophone de la Communication Parlée (AFCP), tentent depuis plusieurs années d'attirer l'attention des pouvoirs publics, des magistrats, des médias et des citoyens sur les limites scientifiques et technologiques des techniques et des procédés d'identification des individus par leur voix (authentification vocale). Cette position a donné lieu à plusieurs textes [1,2] et articles scientifiques [3, 4] ainsi qu'à plusieurs interventions auprès des tribunaux. Ce communiqué vise à réaffirmer et à diffuser cette prise de position.

« Empreintes vocales », un terme inadapté

Le terme « d'empreintes vocales » fréquemment utilisé dans le langage courant et dans les œuvres de fiction donne à croire que la voix présente des caractéristiques tout aussi fiables que celles des empreintes digitales ou génétiques : **il n'en est rien !** A ce jour, les recherches scientifiques ne permettent pas de dire que la voix possède des caractéristiques permettant d'identifier de façon unique chaque être humain.

La voix n'est pas une empreinte digitale ou génétique

La voix présente des différences majeures avec les empreintes digitales et génétiques :

- La voix évolue au cours du temps, que ce soit à court terme (moment de la journée), à moyen terme (période de l'année) et à long terme (avec l'âge), ainsi qu'en fonction de l'état de santé ou l'état émotionnel.
- La voix est un élément modifiable volontairement (cf. les imitateurs) et aisément falsifiable, avec les moyens techniques existants.

De plus, l'évaluation scientifique de la fiabilité des empreintes digitales et génétiques repose notamment sur l'existence de bases de données expérimentales de dimension très importante. Dans le domaine vocal, les bases de données disponibles actuellement ne comportent pas un nombre suffisant de locuteurs, de langues, de conditions d'enregistrement pour évaluer la fiabilité des méthodes existantes dans un contexte d'authentification vocale.

« Identification vocale », des procédés non étayés par une démarche scientifique

Dans l'état actuel des connaissances, il n'existe pas de procédures, automatiques ou basées sur une « expertise », permettant d'avancer avec certitude qu'une personne est - ou n'est pas - l'auteur d'un enregistrement vocal [1].

Ceci est d'autant plus vrai lorsqu'il s'agit d'authentifier un enregistrement « de durée limitée, avec fond sonore important, enregistré dans de mauvaises conditions techniques et provenant d'un locuteur ayant pu déguiser ou modifier artificiellement sa voix » [2].

Des travaux internationalement reconnus, effectués avec la plus grande rigueur méthodologique, tels que ceux qui sont publiés dans les revues internationales confirment scientifiquement ces affirmations [5] [6].

¹ L'AFCP a été créée à l'initiative du Groupe Francophone de la Communication Parlée (GFCP) de la Société Française d'Acoustique (SFA) en novembre 2001.

La « reconnaissance automatique du locuteur », une technologie imparfaite utilisable dans certain cas

La voix contient des informations caractérisant partiellement un locuteur. Le domaine scientifique utilisant ces informations pour vérifier l'identité d'une personne est appelé « reconnaissance automatique du locuteur ». Les applications habituellement visées concernent l'authentification de l'utilisateur, que ce soit pour un accès matériel (à un local, par exemple) ou à un service (notamment, par l'intermédiaire du téléphone).

Les techniques de reconnaissance automatique du locuteur sont basées sur la mesure de ressemblances entre des enregistrements de parole. Ces mesures sont faites sur des paramètres acoustiques extraits par analyse du signal. Elles prennent en compte des caractéristiques statistiques du locuteur, ainsi que le contenu du message vocal, les informations sur l'environnement et le matériel d'enregistrement.

Pour garantir un niveau de performance acceptable dans les contextes applicatifs habituels, les conditions suivantes doivent être remplies :

- Le locuteur ne cherche pas à masquer sa voix : il cherche à être reconnu et il coopère avec le système.
- Les conditions d'enregistrement et de traitement du signal audio sont connues et/ou contrôlées.
- La mesure de ressemblance est préalablement étalonnée au cours d'expériences réalisées dans des conditions comparables au contexte d'utilisation. La décision doit être calibrée en fonction des résultats de ces expériences et en fonction de l'application visée².
- Les imposteurs éventuels sont dissuadés d'avoir recours à des moyens technologiques sophistiqués permettant de falsifier une voix. Ceci est en général induit par l'ergonomie du système.
- En outre, il est préférable que le contenu linguistique du message soit composé de mots déjà connus par le système, pour que la ressemblance entre les voix puisse être calculée sur des éléments comparables.

Sans coopération du locuteur, sans contrôle du matériel de prise de son, sans connaissance des conditions d'enregistrement, sans certitude qu'il ne s'agit pas d'une voix falsifiée et - dans une moindre mesure - sans contrôle du contenu linguistique des messages, il n'est pas possible de déterminer à quel point une ressemblance entre deux messages vocaux est due au locuteur et non à d'autres facteurs.

A ce jour, il n'existe pas de procédé scientifique « d'identification d'une personne par sa voix »

Références bibliographiques

- [1] Pétition pour l'arrêt des expertises vocales, tant qu'elles n'auront pas été validées scientifiquement. Pétition du GFCP de la SFA, 1999. (<http://www.afcp-parole.org/doc/petition.pdf>)
- [2] Motion adoptée à l'unanimité par le Bureau du GCP (Groupe de la Communication Parlée) de la SFA, reconduite intégralement par le GFCP de la SFA en 1997 et par l'AFCP en 2002. (http://www.afcp-parole.org/doc/MOTION_1990.pdf).
- [3] L. J. Boë, F. Bimbot, J.F. Bonastre, P. Dupont, De l'évaluation des systèmes de vérification du locuteur à la mise en cause des expertises vocales en identification juridique, *Langues*, Vol. 2, n°4 Décembre 1999, pp 270-288 (<http://www.afcp-parole.org/doc/Article-Langue.pdf>)
- [4] L.J. Boë, Forensic voice identification in France, *Speech Communication*, ed. Elsevier, Volume 31, Issues 2-3, June 2000, pp. 205-224
- [5] C. Champod, D. Meuwly, The Inference of identity in forensic speaker identification, *Speech Communication*, ed. Elsevier, Volume 31 (2000), pp. 193-203 (<http://www.unil.ch/ipsc/pdf/science.pdf>)
- [6] Martin, A., Przybocki, M., "The NIST 1999 Speaker Recognition Evaluation - An Overview", *Digital Signal Processing*, Vol. 10, Num. 1-3. January/April/July 2000, pp. 1-18

² Pour une même mesure, dans un même environnement, la réponse à la question « est-ce la même personne qui a prononcé ces deux messages vocaux ? » peut être différente suivant que le système privilégie le facteur sécurité (ne pas répondre oui à tort à la question précédente) ou la satisfaction du client (ne pas répondre non à tort à la question précédente).